

# Resolução livro

Grupo: Eduardo Garcez, Davi Augusto, Gabriel Netto, João Arend e Luan Frederico.

## Contents

9.3 Calibração de doses de radiação - Regressão Poisson . . . . .	1
---	---

### 9.3 Calibração de doses de radiação - Regressão Poisson

#### 9.3.1 Objectivo do estudo e introdução ao problema

Depois da exposição accidental quer a valores baixos ou elevados de radiação, é importante poder dar informação precisa sobre a dose de radiação a que vítimas do acidente foram expostas.

Uma informação inicial (a priori) sobre esta dose de radiação pode ser obtida através de dosímetros pessoais e através de avaliação médica qualificada. A quantidade de incerteza associada a esta avaliação é porém elevada.

Para reduzir essa quantidade de incerteza, Groer e Pereira (1986) propõem um método que permite obter uma densidade de calibração da dose de radiação desconhecida a que uma vítima foi exposta, combinando aquela informação a priori com informação sobre o número de aberrações cromossómicas observadas numa amostra de  $n$  células da vítima pouco tempo decorrido após a exposição e com dados obtidos em experiências controladas in vitro. Tipicamente, em experiências desta natureza,  $n_i$  células são expostas a doses de radiação  $d_i$ , rigorosamente controladas pelo experimentador, observando-se ao microscópio o número de aberrações cromossómicas  $y_i$  resultantes. A densidade de calibração da dose de radiação desconhecida  $d_f$ . (onde o índice é relativo a um futuro indivíduo acidentado), assim obtida caracteriza a incerteza acerca da dose de radiação após consideração da informação a priori e dos dados existentes, permitindo ao médico decidir se a vítima em questão necessita ou não de tratamento.

De acordo com a literatura especializada, Groer e Pereira (1986) sugerem utilizar para  $Y$  – o número de aberrações cromossómicas induzidas por uma dose de radiação conhecida  $d$  numa amostra de células – uma distribuição de Poisson de valor médio  $\mu = \alpha nd$ , onde  $\alpha$  é um parâmetro desconhecido. Usando distribuições a priori Gama para  $\alpha$  e para  $d_f$  obtém uma forma analítica para a densidade de calibração da dose de radiação a que uma vítima foi sujeita, após incorporação de informação sobre o número  $y$  de aberrações cromossómicas observadas no indivíduo numa amostra de  $n_y$  células e dos dados de uma experiência in vitro.

Este modelo, tal como eles observam, não é adequado para doses baixas de radiação. O modelo  $Y \sim Poi[n(\alpha_0 + \alpha d + \beta d^2)]$  é outro modelo sugerido na literatura especializada, principalmente para dados relativos a doses de radiação baixa. No entanto para este modelo já não é possível encontrar uma forma analítica para a densidade de calibração e portanto já não é desenvolvido nesse trabalho.

Nesta secção ir-se-á refazer a análise de Groer e Pereira utilizando para  $Y$  um modelo de regressão de Poisson sugerido em Leonard (2000), isto é, admitindo que  $Y \sim Poi(\mu)$  com  $\log(\mu) = \beta_0 + \beta_1 \log(n) + \beta_2 \log(d)$ . Este modelo corresponde ao modelo de Groer e Pereira quando  $\beta_1 = \beta_2 = 1$ . Será ainda analisado o modelo  $Y \sim Poi[n(\alpha_0 + \alpha d + \beta d^2)]$ .

#### 9.3.2 O problema de calibração do ponto de vista bayesiano

O problema de calibração ou calibragem é normalmente entendido como um processo de ajustamento da escala de um aparelho de medição à custa de uma experiência informativa (experiência de calibração) realizada

para esse efeito. No entanto, o problema pode ser formalmente encarado com uma maior generalidade.

Suponha-se, por exemplo, que o resultado  $z$  de uma experiência está relacionado com o valor  $t$  de uma variável controlada através de um modelo de regressão linear simples. Uma experiência informativa resulta em  $n$  pares  $(x_i, t_i)$ . O objectivo do problema de calibração é o de saber qual o valor de  $t_y$  que deu origem a um valor observado  $z_y$  em uma experiência futura (aqui o futuro é entendido num sentido lato, pois pode ter acontecido, temporalmente, antes da própria experiência informativa). Neste caso tem-se um problema de regressão inversa.

O problema de calibração é tratado, do ponto de vista bayesiano, por Aitchison e Dunsmore (1975).

Para formalizar o problema de calibração, considere-se então que a cada unidade experimental está associado um valor  $t$  de um conjunto bem especificado  $T$ . A este valor designa-se por índice da unidade experimental. No exemplo de regressão o índice será o valor da variável controlada. O objectivo do problema de calibração é o de identificar, o melhor possível, os índices desconhecidos de unidades experimentais.

Designa-se por  $z = (x, t) = \{(x_1, t_1), \dots, (x_n, t_n)\}$  o resultado de uma experiência de calibração com a unidades experimentais, onde se identificou, para cada unidade,  $z$  e o índice  $t$ . Suponha-se que são válidas as seguintes hipóteses:

- A distribuição do vector  $Z$  é  $f(t, x|\omega) = f(t|\psi)f(x|t, \theta)$  onde o espaço paramétrico se decompõe no produto cartesiano de dois espaços paramétricos, isto é,  $\omega = (\psi, \theta) \in \Omega = \Psi \times \Theta$
- Os parâmetros  $\psi$  e  $\theta$  são independentes a priori e portanto a distribuição a priori para  $h(\omega)$  pode escrever-se na forma  $h(\omega) = h(\psi)h(\theta)$ .
- Unidades experimentais futuras seguem o mesmo padrão, isto é, se  $u$  designar o índice de uma unidade experimental futura para a qual se observou  $y$  tem-se que  $f(u|\omega) = f(u|\psi)e f(y|t, \omega) = f(y|t, \theta)$ .

Assim tem-se que:

$$h(u, \psi, \theta|y, z) \propto f(y, z|u, \psi, \theta)h(u, \psi, \theta) \propto f(t|\psi)f(x|t, \theta)f(y|u, \theta)f(u|\psi)h(\theta)h(\psi) \propto f(u|\psi)h(\psi|t)h(\theta|z)f(y|u, \theta)$$

e portanto a distribuição de calibração é

$$p(u|y, x) \propto \int \int f(u|\psi)h(\psi|t)d\psi f(y|u, \theta)h(\theta|x)d\theta$$

(9.10)

$$\propto p(u|t)p(y|u, z)$$

(9.11)

O problema de calibração pode assim ser encarado como um problema inverso da predição, embora a distribuição preditiva desempenhe um papel predominante.

Voltando ao exemplo apresentado na introdução, designa-se por  $D = \{(y_i, n_i, d_i), i = 1, \dots, m\}$  o conjunto de todos os dados resultantes de uma experiência in vitro em que  $n_i$  células são expostas a uma dose conhecida de radiação  $d_i$ , tendo-se observado  $y_i$  aberrações. Seja ainda  $y_f, n_f$  os dados obtidos de uma vítima de exposição a uma dose  $d_f$  desconhecida de radiação. Suponha-se que  $Y_i$   $i = 1, \dots, m$  são mutuamente independentes com distribuição  $f(y_i|n_i, d_i, \beta)$  onde é um vector de parâmetros desconhecido, Admita-se ainda que, dado  $\beta$ ,  $Y_f$  é independente dos  $Y_i$  e  $d_f$  é independente de  $D$ .

A densidade de calibração de  $d_f$  é assim:

$$p(d_f|y_f, n_f, \mathcal{D}) = \int h(d_f, \beta|y_f, n_f, \mathcal{D})d\beta = \int h(d_f|\beta, y_f, n_f, \mathcal{D})h(\beta|y_f, n_f, \mathcal{D})d\beta = \int \frac{f(y_f|n_f, d_f, \beta)h(d_f|\beta)}{f(y_f|n_f, \beta)} \frac{f(y_f|n_f, \beta)h(\beta|\mathcal{D})}{f(y_f|n_f, \mathcal{D})}d\beta \\ = \int \frac{f(y_f|n_f, d_f, \beta)h(d_f)}{f(y_f|n_f, \beta)} \frac{f(y_f|n_f, \beta)h(\beta|\mathcal{D})}{p(y_f|n_f, \mathcal{D})}d\beta \propto h(d_f)p(y_f|n_f, d_f, \mathcal{D})$$

onde  $h(d_f)$  é a distribuição a priori para  $d_f$  a qual se supõe ser independente de  $\beta$ . Esta equação estabelece que a densidade de calibração de  $d_f$  é proporcional à distribuição a priori para  $d_f$  e à densidade preditiva de um número futuro de aberrações cromossómicas resultantes. Note-se que ela é idêntica a (9.11). A diferença reside no facto de neste caso as doses de radiação  $d_i$  experimentais serem fixas e não aleatórias.

Supondo que  $Y \sim Poi(\alpha nd)$ ,  $\alpha \sim Ga(a, b)$ ,  $d_f \sim Ga(c, e)$ , é fácil de mostrar que a distribuição preditiva de  $Y_f$  é binomial negativa, isto é

$$p(y_f|n_f, d_f, \mathcal{D}) = \binom{y_f+A-1}{A-1} \left(\frac{B}{n_f d_f + B}\right)^A \left(\frac{n_f d_f}{n_f d_f + B}\right)^{y_f}, y_f = 0, 1, \dots,$$

sendo o valor médio e variancia, respectivamente, dados por

$$E[Y_f|n_f, d_f, \mathcal{D}] = \frac{An_f d_f}{B}$$

$$var[Y_f|n_f, d_f, \mathcal{D}] = \frac{An_f d_f (n_f d_f + B)}{(B)^2}$$

e que

$$p(d_f|y_f, n_f, \mathcal{D}) \propto d_f^{C-1} (d_f + E)^{-(A+y_f)} \exp(-Bd_f),$$

onde  $C = c + y_f$ ,  $E = \frac{e + \sum_{i=1}^m n_i d_i}{n_f}$ ,  $A = \sum_{i=1}^m y_i + a$  e  $B = b + \sum_{i=1}^m n_i d_i$ ,

### 9.3.3 Aplicação aos dados

Nesta secção ir-se-á aplicar a teoria descrita considerando:

- (i) o modelo de regressão de Poisson previamente descrito para o número de aberrações cromossómicas, isto é  $Y \sim Poi(\mu)$ ,  $\log(\mu) = \beta_0 + \beta_1 \log(n) + \beta_2 \log(d)$ ,
- (ii) os modelos propostos em Groer e Pereira (1986) nomeadamente,  $Y \sim Poi(\mu)$ ,  $\mu = \alpha nd$

**Tabela 9.9: Dados de radiação por neutrões Po-Be.**

i	Nº de aberrações	Nº de células (1000) $n_i$	nível de dose $d_i(rad)$
1	109	269	50
2	47	78	100
3	94	115	75
4	114	90	200
5	138	84	150
6	125	59	250
7	97	37	300
f	64 / 13	104	$d_{f_2} / d_{f_3}$

**Tabela 9.10: Dados de radiação - The Lundsteen-Piper (1989) lymphocyte data.**

i	Nº de aberrações	Nº de células (1000) $n_i$	nível de dose $d_i(rad)$
1	0	585	0.10
2	3	1002	0.20
3	5	472	0.50
4	14	493	1.00
5	30	408	1.50
6	75	690	2.00
7	46	291	3.00
f	20	700	$d_{f_1}$

- (iii)  $Y \sim Poi(\mu)$ ,  $\mu = n(\alpha_0 + \alpha_1 d + \alpha_2 d^2)$ .

Os dados utilizados encontram-se nas Tabelas 9.9 e 9.10. As últimas linhas das tabelas dizem respeito a observações de indivíduos futuros para os quais se pretende obter uma densidade de calibração para a dose de radiação desconhecida a que foram expostos.

Começar-se-á com o estudo da adequabilidade dos modelos em estudo e consequente selecção do melhor modelo para cada conjunto de dados, utilizando para tal as técnicas sugeridas na Secção 8.4.

Dado que não é possível obter, para os modelos em consideração (i) e (iii), expressões analíticas quer para a distribuição preditiva de  $y_f$ , quer para a densidade de calibração de  $d_f$ , utilizou-se o software WinBUGS para o efeito.

Para a obtenção dos resíduos de eliminação

$$d_{1i} = y_{i.obs} - E[Y_i|y_{(-i)}]$$

$$d_i^f = \frac{y_{i.obs} - E[Y_i|y_{(-i)}]}{\sqrt{var[Y_i|y_{(-i)}]}}$$

e das ordenadas preditivas condicionais (CPO)

$$p(y_i|y_{(-i)}) = \int f(y_i|\theta, y_{(-i)})h(\theta|y_{(-i)})d\theta$$

em que denota o vector genérico dos parâmetros de cada modelo, os cálculos necessários foram feitos retirando do modelo uma observação de cada vez, já que a dimensão da amostra o permitia.

Para os modelos (i) e (iii), em que se utilizou o software WinBUGS, e dado que se  $Y \sim Poi(\mu)$ ,  $E(Y|\mu) = \mu$ , as ordenadas preditivas condicionais, os valores médios preditivos e as variancias foram calculadas usando

$$p(y_{i.obs}|y_{(-i)}) \approx \frac{1}{n} \sum_{j=1}^n \exp(-\mu_i^*) \frac{(\mu_i^*)^{y_{i.obs}}}{y_{i.obs}!}. \quad E[Y_i|y_{(-i)}] \approx \frac{1}{n} \sum_{j=1}^n \mu_i^* \quad E[Y_i^2|y_{(-i)}] \approx \frac{1}{n} \sum_{j=1}^n \mu_i^* + (\mu_i^*)^2,$$

$$var[Y_i|y_{(-i)}] = E[Y_i^2|y_{(-i)}] - (E[Y_i|y_{(-i)}])^2,$$

com base numa amostra  $\{\theta_j^*, j = 1, \dots, n\}$  n da distribuição a posteriori e  $\mu_i^* = g(\theta_i^*)$  de acordo com o modelo considerado.

### Informação a priori

Para cada um dos modelos em estudo foi considerada a seguinte informação a priori

**Modelo (i)**  $Y \sim Poi(\mu)$ ,  $\log(\mu) = \beta_0 + \beta_1 \log(n) + \beta_2 \log(d)$  Admitiu-se que a priori os parametros do modelo  $\beta_0, \beta_1, \beta_2$  eram independentes e normalmente distribuídos com valores médios zero e variancias elevadas (10000), o que corresponde a uma informação vaga. Para a distribuição a priori de  $d_f$  para os dados da Tabela 9.9, seguiu-se a sugestão de Groer e Pereira, nomeadamente,  $d_f \sim Ga(10, 0.1)$ .

**Modelo (ii)**  $Y \sim Poi(\mu)$   $\mu = \alpha n d$  Para este modelo considerou-se para o uma distribuição a priori Gama,  $Ga(a, b)$ . Duas situações foram estudadas; informação a priori vaga, isto é  $a = b = 0$  e a informação a priori sugerida por Groer e Pereira, nomeadamente  $a = 10$ ,  $b = 1000$ . Para a distribuição a priori de  $d_f$  para os dados da Tabela 9.10, seguiu-se novamente a sugestão de Groer e Pereira, isto é,  $d_f \sim Ga(10, 0.1)$ .

**Modelo (iii)**  $Y \sim Poi(\mu)$ ,  $\mu = n(\alpha_0 + \alpha_1 d + \alpha_2 d^2)$ . A escolha de uma distribuição a priori para os parâmetros deste modelo requer cuidados especiais. Em princípio, o espaço de parâmetros deste modelo é  $R^3$ , tal como para o modelo de regressão Poisson e, portanto, se se quiser admitir, por exemplo, uma distribuição a priori não informativa, poder-se-á considerar que os parâmetros são independentes a priori e uniformes em IR. No entanto, os parâmetros estão sujeitos à restrição  $\alpha_0 + \alpha_1 d + \alpha_2 d^2 > 0$ , o que traz problemas adicionais, principalmente se não se esperar para a valores consideravelmente elevados, que é o que acontece com o segundo conjunto de dados. Um modo de ultrapassar o problema da restrição imposta aos parâmetros pelo facto de  $\mu > 0$  é admitir que parâmetros são todos positivos. Esta hipótese é, por ela própria, também restritiva.

Tabela 9.11: Estimativas pontuais e intervalos de credibilidade 95% para os parâmetros dos modelos

Modelo	Parâmetro	média	desvio padrão	2.5%	mediana
(i)	$\beta_0$	-4.752	1.39	-7.702	-4.701
	$\beta_1$	0.9763	0.1549	0.6876	0.9719
	$\beta_2$	1.014	0.1477	0.7343	1.009
(ii-a)	$\alpha$	0.008327	0.000309	0.007720	0.008327
(ii-b)	$\alpha$	0.008346	0.000308	0.007742	0.008346
(iii-a)	$\alpha_0$	0.01092	0.1202	-0.2093	0.01343
	$\alpha_1$	0.007805	0.002366	0.003093	0.007676
	$\alpha_2$	2.80E-06	8.29E-06	-1.25E-05	3.15E-06
(iii-b)	$\alpha_0$	0.09803	0.07159	0.005422	0.08338
	$\alpha_1$	0.006112	0.001407	0.002699	0.006365
	$\alpha_2$	8.43E-06	5.22E-06	6.54E-07	7.69E-06
(iii-c)	$\alpha_0$	0.08225	0.06138	0.003541	0.07004
	$\alpha_1$	0.006361	0.001225	0.003612	0.006523
	$\alpha_2$	7.72E-06	4.67E-06	5.91E-07	7.17E-06

\*Modelo (i): modelo de Regressão de Poisson;  
 Modelo (ii-a): modelo de Groer e Pereira com distribuição a priori vaga, isto é,  $a = 1$ ,  $b = 1$ ;  
 Modelo (ii-b): mesmo modelo com informação a priori  $Ga(10, 1000)$ ;  
 Modelo (iii-a): modelo (iii) com distribuição a priori normal, usando truncatura;  
 Modelo (iii-b): modelo (iii) com distribuição a priori  $Ga(1, 1)$ ;  
 Modelo (iii-c): modelo (iii) com distribuição a priori  $Ga(1, 4)$ .

elos - Dados da Tabela 9.9.

Assim consideraram-se várias situações de informação a priori:

- Admitiu-se que a priori os parâmetros do modelo  $\alpha_0$  e  $\alpha_2$  eram independentes e normalmente distribuídos com valores médios zero e variâncias elevadas (10000), o que corresponde a uma informação vaga. Usou-se truncatura para  $\alpha_1$ , isto é, sempre que era gerado um valor de  $\mu < 0$  ele era rejeitado.
- Para o primeiro conjunto de dados isto não representou problema, pois os valores de gerados eram sempre positivos. O mesmo já não aconteceu para o segundo conjunto de dados.
- Admitiu-se que a priori os parâmetros do modelo  $\alpha_0$  e  $\alpha_2$  eram independentes com distribuição Gama  $Ga(a, b)$ . Consideraram-se vários valores para os hiperparâmetros, tendo-se optado por restringir o estudo a duas situações  $a = b = 1$ ,  $a = 1, b = 4$

### Adequabilidade e comparação de Modelos

Como se referiu, antes de considerar o problema da calibração e de estimar a dose de radiação a que indivíduos futuros foram sujeitos, ir-se-á, para cada conjunto de dados, estudar a adequabilidade dos vários modelos em estudo e a selecção do melhor modelo.

Inicialmente ajustaram-se os modelos a cada conjunto de  $m = 7$  dados. Nas Tabelas 9.11 e 9.12 apresenta-se um resumo do output do WinBUGS sobre as estatísticas relativas aos parâmetros dos modelos ajustados para os dois conjuntos de dados. Os resultados em relação ao modelo (ii) são exactos, isto é, foram obtidos usando a distribuição a posteriori Gama,  $Ga(A, B)$ , para  $\alpha$ .

Tabela 9.12: Estimativas pontuais e intervalos de credibilidade 95% para os parâmetros dos modelos - Dados da Tabela 9.10.

Modelo	Parâmetro	média	desvio padrão	2.5%	mediana	97.5%
(i)	$\beta_0$	-5.196	1.441	-8.031	-5.184	-2.4
	$\beta_1$	1.282	0.2248	0.8425	1.281	1.725
	$\beta_2$	1.638	0.1665	1.328	1.635	1.973
(ii-a)	$\alpha$	0.044901	0.003414	0.038210	0.044901	0.051592
(ii-b)	$\alpha$	0.037709	0.002788	0.032246	0.037709	0.043173
(iii-a)	$\alpha_0$	-0.004058	0.002652	-0.009358	-0.003989	1.11E-03
	$\alpha_1$	0.03266	0.01276	0.008598	0.03252	0.05873
	$\alpha_2$	0.009432	0.005776	-0.001981	0.009382	0.02064
(iii-b)	$\alpha_0$	0.001128	0.001041	3.12E-05	8.33E-04	0.003824
	$\alpha_1$	0.01309	0.00715	0.001317	0.01246	0.02882
	$\alpha_2$	0.01742	0.004023	0.008996	0.01762	0.02458
(iii-c)	$\alpha_0$	0.00114	0.001038	3.27E-05	8.43E-04	0.003831
	$\alpha_1$	0.01292	0.007111	0.001256	0.01234	0.02825
	$\alpha_2$	0.01749	0.004002	0.00922	0.01767	0.02469

\*Modelo (i) é o modelo de Regressão de Poisson,  
Modelo (ii-a) é o modelo de Groer e Pereira com distribuição a priori vaga, isto é,  $a = b = 0$ ,  
Modelo (ii-b) é o mesmo modelo com informação a priori  $Ga(10, 1000)$ ,  
Modelo (iii-a) é o modelo (iii) com distribuição a priori normal, usando truncatura,  
Modelo (iii-b) é o modelo (iii) com distribuição a priori  $Ga(1, 1)$ ,  
Modelo (iii-c) é o modelo (iii) com distribuição a priori  $Ga(1, 4)$ .

Através de uma análise preliminar destes resultados, pode-se pensar que, no que diz respeito ao modelo (i), a hipótese  $\beta_1 = \beta_2 = 1$  (os modelos (i) e (ii) são idênticos) não será de rejeitar e como tal o modelo (ii) será preferível, já que tem menos parâmetros. Relativamente ao modelo (iii-a), verifica-se também que, possivelmente, não será de rejeitar a hipótese de  $\alpha_0 = \alpha_2 = 0$ , o que reduz novamente este modelo ao modelo (ii). Devido às dificuldades associadas ao seu estudo não se prosseguiu com a sua análise, preferindo continuar com os modelos (iii-b) e (iii-c).

Novamente através de uma análise preliminar destes resultados pode-se conjecturar que, nesta situação, os modelos (i) e (ii) são distintos. Com efeito, as estatísticas referentes ao parâmetro  $\beta_2$  sugerem que um modelo adequado deve incluir uma potência de  $d$  maior que 1, mas não necessariamente 2, como admite o modelo (iii). Note-se agora uma diferença entre os modelos (ii-a) e (ii-b). Deve referir-se que a informação a priori sugerida em Groer e Pereira para  $\alpha$  tinha sido eliciada para experiências com doses de radiação elevadas e, portanto, essa informação não é necessariamente a mais adequada nesta situação. Relativamente ao modelo (iii-a) parecem ser válidas conclusões semelhantes às tiradas com os dados anteriores. Haverá necessidade de efectuar uma análise mais cuidada de como lidar com a restrição imposta aos parâmetros, pois o método de truncatura utilizado pode não ser o mais adequado (veja-se, e.g., Chen et al., 2000, cap. 6). Não parece haver diferenças significativas entre os modelos (iii-b) e (iii-c), aliás como seria de esperar.

Para se fazer uma análise mais detalhada da adequabilidade dos diferentes modelos e seleccionar o melhor, reajustaram-se os diferentes modelos aos dados após retirar um caso de cada vez, permitindo assim obter uma avaliação mais precisa dos erros bayesianos de eliminação e das ordenadas preditivas condicionais, instrumentos que serão utilizados nessa análise,

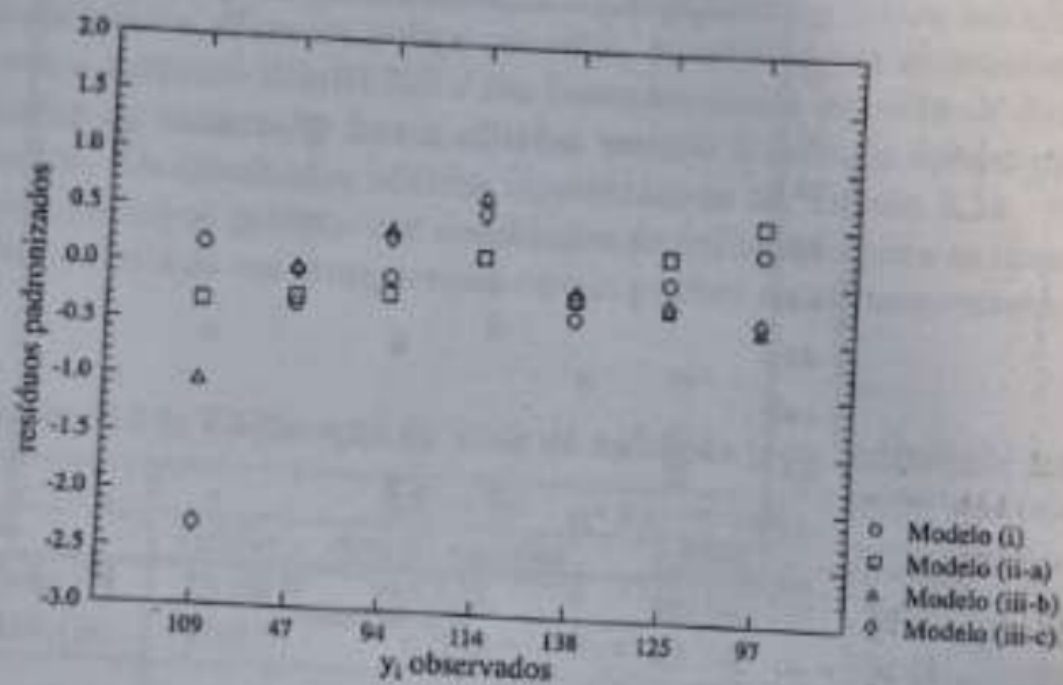
Através de uma análise preliminar destes resultados, pode-se pensar que, no que diz respeito ao modelo (i), a hipótese  $\beta_1 = \beta_2 = 1$  (os modelos (i) e (ii) são idênticos) não será de rejeitar e como tal o modelo (ii) será preferível, já que tem menos parâmetros. Relativamente ao modelo (iii-a), verifica-se também que, possivelmente, não será de rejeitar a hipótese de  $\alpha_0 = \alpha_2 = 0$ , o que reduz novamente este modelo ao modelo (ii). Devido às dificuldades associadas ao seu estudo não se prosseguiu com a sua análise, preferindo continuar com os modelos (iii-b) e (iii-c).

Novamente através de uma análise preliminar destes resultados pode-se conjecturar que, nesta situação, os modelos (i) e (ii) são distintos. Com efeito, as estatísticas referentes ao parâmetro  $\beta_2$  sugerem que um modelo adequado deve incluir uma potência de  $d$  maior que 1, mas não necessariamente 2, como admite o modelo (iii). Note-se agora uma diferença entre os modelos (ii-a) e (ii-b). Deve referir-se que a informação a priori sugerida em Groer e Pereira para  $a$  tinha sido eliciada para experiências com doses de radiação elevadas e, portanto, essa informação não é necessariamente a mais adequada nesta situação. Relativamente ao modelo (iii-a) parecem ser válidas conclusões semelhantes às tiradas com os dados anteriores. Haverá necessidade de efectuar uma análise mais cuidada de como lidar com a restrição imposta aos parâmetros, pois o método de truncatura utilizado pode não ser o mais adequado (veja-se, e.g., Chen et al., 2000, cap. 6). Não parece haver diferenças significativas entre os modelos (iii-b) e (iii-c), aliás como seria de esperar.

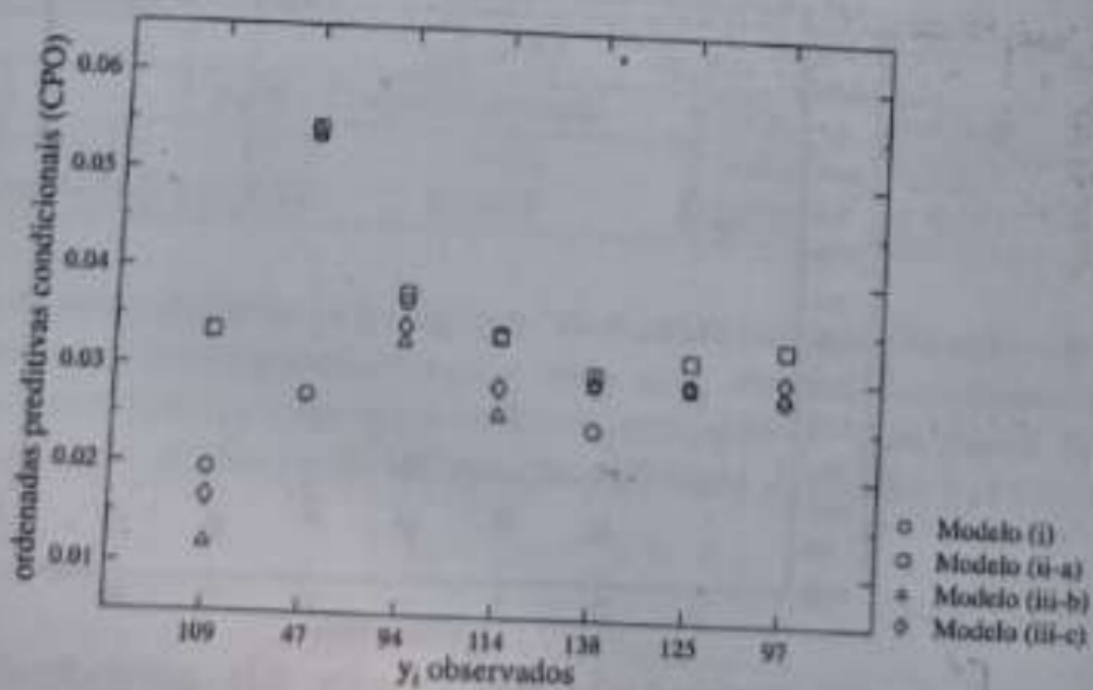
Para se fazer uma análise mais detalhada da adequabilidade dos diferentes modelos e seleccionar o melhor, reajustaram-se os diferentes modelos aos dados após retirar um caso de cada vez, permitindo assim obter uma avaliação mais precisa dos erros bayesianos de eliminação e das ordenadas preditivas condicionais, instrumentos que serão utilizados nessa análise,

Nas Figuras 9.4(a,b) e 9.5(a,b) apresentam-se gráficos, para os dados referentes às Tabelas 9.9 e 9.10 respectivamente, dos resíduos bayesianos padronizados e das ordenadas preditivas condicionais para os modelos em competição.

Na Tabela 9.13 encontram-se medidas sumárias dessas quantidades, nomeadamente, a soma dos quadrados dos resíduos padronizados (SQRS) e a soma dos logaritmos das ordenadas preditivas condicionais (SinCPO), tal como é sugerido na Secção 8.4.2. Um bom modelo deverá ter um valor baixo de SQRS e elevado de (SinCPO).



(a) resíduos de eliminação bayesianos padronizados;



(b) ordenadas preditivas condicionais.

(a) resíduos de eliminação bayesianos padronizados;



(b) ordenadas preditivas condicionais.

Pela análise da Figura 9.4, observa-se que os modelos parecem ser todos adequados, embora a observação correspondente a  $y_1 = 109$  não se encontre tão bem ajustada pelos modelos (i) e (iii). A observação correspondente a  $y_2 = 47$  também não é bem ajustada pelo modelo (i). A observação da Figura 9.5 revela um padrão de resíduos estranho para todos os modelos, o que faz suspeitar da não adequabilidade da suposição de uma distribuição de Poisson para modelar o número de aberrações.